Our team aims at movie quality prediction with the dataset of over 5000 movies from
`https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset`.

Our work is meaningful in the era of fast-developing movie industry. On one hand, audiences can evaluate movies via their public information like directors and posters, so as to select and enjoy good movies. On the other hand, movie makers get to know and design the proper elements for successful movies, for example, what topics fit the trend, which director and actors are the best match to a certain genre of movie, etc.

The dataset is suitable to our project, since it has quality measures of 5000+ movies, including the return rate (Gross/Budget), the imdb score, as well as their other public features. For each quality measure, we can consider it as the target variable adopt multivariate regression, which will select important predictors for movie quality prediction and management. In addition, the data includes numerical, categorical and text variables, which is a good hands on exercise for this course.